

Center for Structural Genomics of Infectious Diseases

Contract No.: HHSN272200700057C

Seattle Structural Genomics Center for Infectious Disease

Principal Investigator: Peter J. Myler, Ph.D.

Seattle Biomedical Research Institute

307 Westlake Ave N

Suite 500

Seattle, WA 98109-5219

Tel.: (206) 256-7332

FAX: (206) 256-7229

e-mail: peter.myler@sbri.org

First report of Targets Proposed for Structure Determination

Author: Peter J. Myler

Date: October 28, 2007

Amended: December 1, 2007, Peter J. Myler
January 3, 2008, Peter J. Myler

Target Selection process

For the initial round (batch01a) of Target Selection, we have followed an abbreviated version of the approach indicated on page 43 of the original proposal (see Figure 1, below). Because of the short period elapsed since the contract start date, we have used essentially the central portion of the Target Selection pipeline, without any input from external experts or scientific community, and forgoing any extensive examination of the literature and/or GO annotation. As indicated in the original proposal, we have focused our efforts on three Gram-negative bacterial genera from the Category B priority list, but decided to use only a single representative strain for each genus for the initial round of Target Selection.

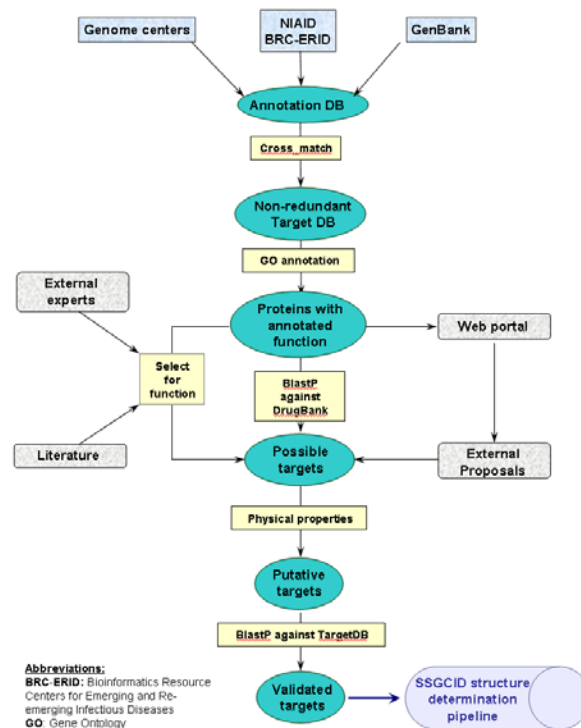


Figure 1: Target selection pipeline for SSGCID

The protein sequences predicted from the following three complete genomes were downloaded from the sequencing center and/or BRC-ERID: *Rickettsia prowazekii* Madrid E (<http://cmr.tigr.org/tigr-scripts/CMR/GenomePage.cgi?org=ntrp01>) , *Brucella melitensis* biovar Abortus (<http://cmr.tigr.org/tigr-scripts/CMR/GenomePage.cgi?org=ntbm02>), and *Burkholderia pseudomallei* 1710a (<http://pathema.tigr.org/tigr-scripts/Burkholderia/GenomePage.cgi?org=gbp1710a>). As shown in Table 1, a total of 11,016 protein sequences were predicted from the three genomes. These were subsequently BLASTED against the DRUGBANK database (downloaded from <http://redpoll.pharmacy.ualberta.ca/drugbank/>), and those sequenced that showed >50% similarity over >75% of their length passed to the next filter. This reduced the number of candidates to 433 (3.9% of the initial number). Table

I shows the number of sequences that passed each step of the process. A series of physical screens were used to eliminate proteins longer than 500 amino acids, containing more than 8 cysteine residues and/or containing any transmembrane spanning domains (predicted using TMPRED and/or TMHMM/PHOBIUS), except for N-terminal signal sequences. In the latter case, the signal sequences will be removed before PCR amplification. The 203 remaining candidate proteins were BLASTED against Target DB (downloaded from <http://www.pdb.org>) and those that showed >95% similarity over >80% of their length eliminated. This ensures that none of the targets has been selected for structure determination at any other center. Only one protein was eliminated at this stage. The remaining protein sequences were BLASTED against each other to determine (with a cut-off of 75% similarity over 75% of their length) whether any represented orthologues or paralogues, and a single representative chosen from each orthologous/paralogous group. This resulted in removal of 5 *Burkholderia* sequences, for a final total of 197 Target Sequences.

Table 1 – Summary of Target Selection screens in batch01

Genus	<i>Rickettsia</i>	<i>Brucella</i>	<i>Burkholderia</i>	Total
Proteins annotated	834	3034	7148	11,016
DrugBank screen	43	140	250	433
<500 amino acids	35	113	204	352
<8 cysteines	34	111	200	345
No predicted TMs	22	66	115	203
TargetDB screen	22	65	115	202
Orthologue screen	22	65	109	196
ftsZ orthology	5	4	4	13

Since only one chromosome from *Burkholderia pseudomallei* 1710a has been submitted to GENBANK, we remapped the 110 proteins from 1710a to 1710b. This resulted in the loss of one sequence which was not present in the latter strain. These proposed targets are summarized in Table 2 and shown in more detail the attached spreadsheet, where they are indicated by our unique ID, as well as the GENBANK ID and putative function (as annotated by the sequencing center and/or BRC-ERID). The SSGCID Target ID has the following format:

GgspA.NNNNN.a

Ggsp represents the first two letters of the genus species names (see Table 3)

A represents a unique alphabetic identifier for the strain

NNNNN represents a unique identifier for the Clustered Orthologous Group (COG) obtained by Jaccard clustering of all possible targets

.a represents a unique identifier for each paralogous gene copy within a given strain

The GENBANK ID for the corresponding nucleotide sequence and coordinates of each CDS within this sequence are also indicated in the spreadsheet.

It is our intention to enter all the selected Targets into the Expression, Crystallization and Structure determination pipeline described in our final proposal (see Figure 2). However, we will deviate from the serial escalation plan by entering the initial 192 targets into Tiers 1, 2 and 3 simultaneously in order to expedite their passage through the pipeline. This will also serve to generate data for a statistically significant comparison of expression and solubility using the bacterial (Tier 1) and wheat germ cell-free translation (Tier 2) systems, as well as comparing the natural (Tiers 1 & 2), with codon-optimized and point-mutated (Tier 3) sequences. We anticipate using X-ray crystallography for structure determination of all Targets that show successful crystallization. We will select ~6 proteins that fail to crystallize and have molecular masses <20 kDa for entry into the NMR pathway (Tier 10).

DRUGBANK represents a comprehensive and publicly available resource that combines detailed drug (*i.e.* chemical) data with comprehensive target (*i.e.* protein) information. The database contains >4,100 drug entries and >14,000 protein sequences that are linked to these drug entries. By selecting bacterial proteins with sequence similarity to known drug targets, we will substantially increase the likelihood that these proteins are “druggable”. In addition, knowledge of chemical ligands (*e.g.* the drugs that act against their DRUGBANK homologues) that are likely to bind these proteins should increase their success in traversing the structure determination pipeline and provide ligands for co-crystallization in Tiers 4, 11 and 12. Determination of their three-dimensional structure will facilitate basic biomedical research by significantly shortening the time needed for development of novel chemotherapeutic agents.

In addition, a second round (batch01b) of target selection was carried out in these three species (and related species within the same genera) based on orthology to the cell division protein *ftsZ* from *Bacillus subtilis*, the structure of which has recently been solved by deCODE. We propose to select one or more of these targets for entry into Tiers 11 (additional ligand screens) and 12 (“Fragments-of-Life” screen) as part of our Research and Development effort.

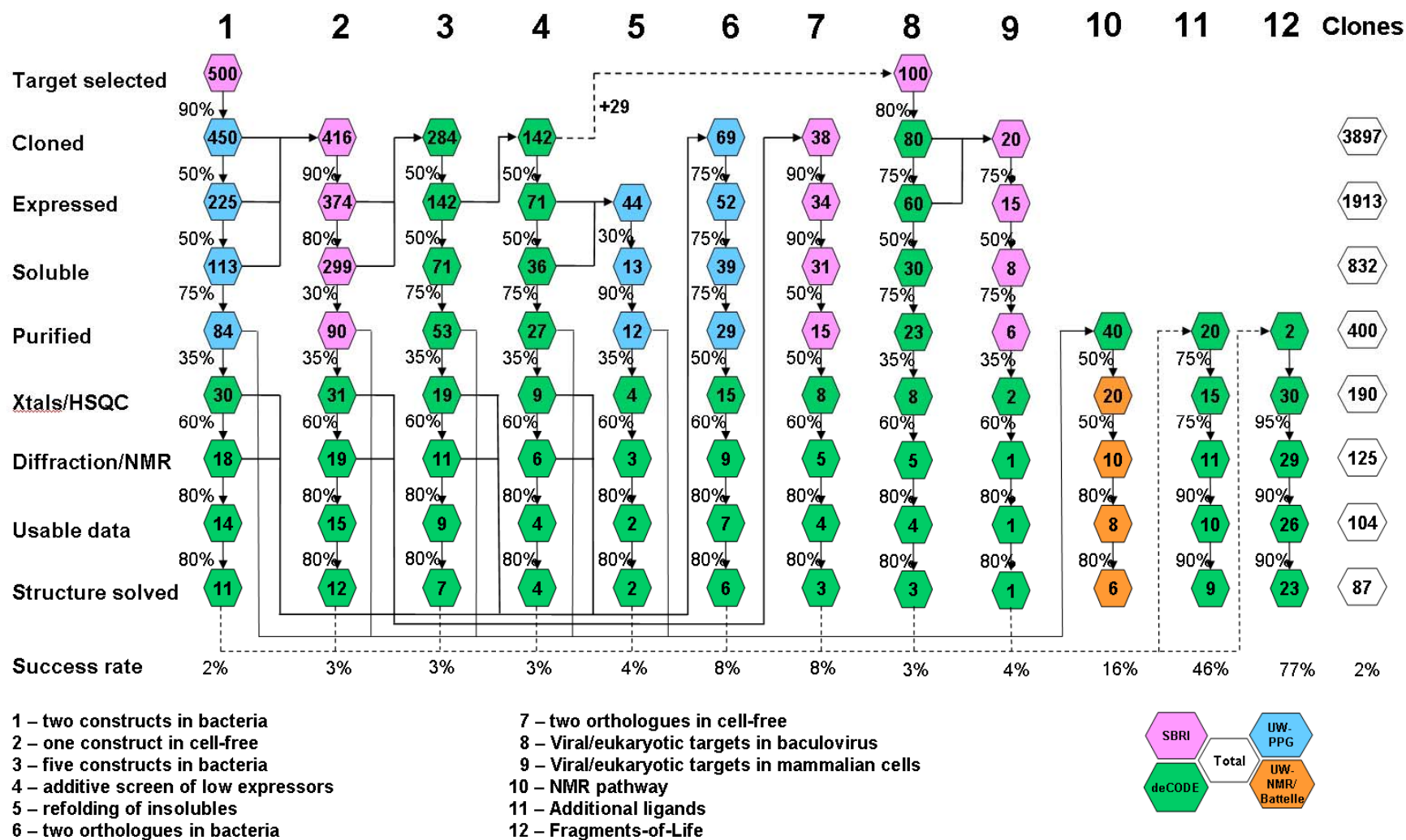


Figure 2: Projected Annual Structure Determination Pipeline.

Table 2 – batch01 Targets selected

batch	target id	gid	refseq id	annotation
1a	BupsA.00001.a	76817596	YP_336223	cytidine deaminase
1a	BrabA.00002.a	83269836	YP_419127	Bacterial transferase hexapeptide repeat:2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase
1a	BupsA.00002.a	76811327	YP_333986	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase
1a	RiprA.00002.a	15604068	NP_220583	2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLATE N-SUCCINYLTRANSFERASE (dapD)
1a	BrabA.00003.a	82700627	YP_415201	Regulator of chromosome condensation, RCC1:Ribosomal protein L27
1a	BupsA.00003.a	76810942	YP_334894	50S ribosomal protein L27
1a	RiprA.00003.a	15604586	NP_221104	50S ribosomal protein L27
1a	BupsA.00004.a	76809418	YP_334791	30S ribosomal protein S9
1a	RiprA.00004.a	15604105	NP_220620	30S ribosomal protein S9
1a	BrabA.00005.a	82700692	YP_415266	Lactate/malate dehydrogenase:L-lactate dehydrogenase:TrkA potassium uptake protein
1a	BupsA.00005.a	76818753	YP_335954	malate dehydrogenase
1a	BrabA.00006.a	82700114	YP_414688	Precorin-8X methylmutase CbiC/CobH
1a	BrabA.00007.a	82699285	YP_413859	Conserved hypothetical protein, ArsC related:Arsonate reductase and related
1a	BrabA.00008.a	82699646	YP_414220	Glycine hydroxymethyltransferase
1a	BupsA.00008.a	76817352	YP_337258	serine hydroxymethyltransferase
1a	RiprA.00008.a	15604577	NP_221095	serine hydroxymethyltransferase
1a	BupsA.00009.a	76810312	YP_332466	kynureninase
1a	BrabA.00010.a	82699333	YP_413907	Short-chain dehydrogenase/reductase SDR:Glucose/ribitol dehydrogenase
1a	BrabA.00010.b	82699369	YP_413943	Short-chain dehydrogenase/reductase SDR:Glucose/ribitol dehydrogenase
1a	BrabA.00010.c	82700918	YP_415492	possible enoyl-(acyl-carrier-protein) reductase
1a	BupsA.00010.a	76808828	YP_334286	3-ketoacyl-(acyl-carrier-protein) reductase
1a	BupsA.00010.b	76811745	YP_334023	enoyl-(acyl carrier protein) reductase
1a	BrabA.00012.a	82699003	YP_413577	Transcriptional coactivator/pterin dehydratase
1a	BrabA.00013.a	82700490	YP_415064	CbxX/CfqX superfamily:Disease resistance protein:ATP/GTP-binding site motif A (P-loop):AAA ATPase:AAA ATPase, central region:...
1a	BupsA.00013.a	76808940	YP_334778	Holliday junction DNA helicase RuvB
1a	RiprA.00013.a	15604251	NP_220767	Holliday junction DNA helicase RuvB
1a	BrabA.00014.a	82700415	YP_414989	Isocitrate lyase and phosphorylmutase:Isocitrate lyase
1a	BupsA.00014.a	76810801	YP_334005	isocitrate lyase
1a	BupsA.00014.b	76819805	YP_336893	methylisocitrate lyase
1a	BrabA.00015.a	82700668	YP_414642	Ribosomal protein S10:Ribosomal protein S10, bacterial form
1a	BupsA.00015.a	76809572	YP_335141	30S ribosomal protein S10
1a	RiprA.00015.a	15604506	NP_221024	30S ribosomal protein S10
1a	BrabA.00016.a	83269774	YP_419065	Blue (type 1) copper domain:Blue (type 1) copper protein:Amicyanin:Plastocyanin
1a	BupsA.00017.a	76811422	YP_332590	cytochrome C552 precursor
1a	BupsA.00018.a	76809671	YP_333972	ribosome releasing factor
1a	BrabA.00019.a	83269116	YP_418407	Aldo/keto reductase
1a	BrabA.00020.a	82700432	YP_415006	Aldehyde dehydrogenase
1a	BupsA.00020.a	76812101	YP_333602	succinate-semialdehyde dehydrogenase (NADP+)
1a	BupsA.00020.b	76817963	YP_335535	betaine aldehyde dehydrogenase
1a	BrabA.00021.a	83269661	YP_418952	Amino acid-binding ACT:D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain:D-isomer specific 2-hydroxyacid dehydr...
1a	BupsA.00022.a	76811593	YP_335153	50S ribosomal protein L1
1a	BrabA.00023.a	82700752	YP_415326	Inorganic pyrophosphatase:Bacterial/Archaeal inorganic pyrophosphatase
1a	BupsA.00023.a	76811279	YP_332648	inorganic pyrophosphatase
1a	RiprA.00023.a	15604440	NP_220958	inorganic pyrophosphatase
1a	BupsA.00024.a	76819379	YP_335782	alpha-ketoglutarate-dependent taurine dioxygenase
1a	BupsA.00025.a	76819159	YP_335674	NAD(+) synthetase
1a	BrabA.00026.a	82699370	YP_413944	Acyl carrier protein (ACP):Phosphopantetheine attachment site:Phosphopantetheine-binding domain
1a	BupsA.00026.a	76809153	YP_334285	acyl carrier protein
1a	BrabA.00027.a	82699928	YP_414502	Acyl-CoA dehydrogenase:Acyl-CoA dehydrogenase, C-terminal:Acyl-CoA dehydrogenase, central domain:Acyl-CoA dehydrogenase, N-te...
1a	BupsA.00027.a	76810110	YP_334609	glutaryl-CoA dehydrogenase
1a	BrabA.00028.a	83269559	YP_418850	Ferritin:Bacterioferritin
1a	BupsA.00028.a	76818131	YP_337087	bacterioferritin
1a	BrabA.00029.a	82700858	YP_415432	Thioredoxin:Thioredoxin type domain:Thioredoxin domain 2
1a	BupsA.00029.a	76812139	YP_333769	thioredoxin 1
1a	RiprA.00029.a	15603883	NP_220398	THIOREDOXIN (trxA)
1a	BrabA.00030.a	82699550	YP_414524	CbxX/CfqX superfamily:ATP/GTP-binding site motif A (P-loop):Bacterial chromosomal replication initiator protein, DnaA:AAA ATP...
1a	BupsA.00030.a	76811706	YP_333873	ATP-dependent protease ATP-binding subunit
1a	RiprA.00030.a	15604535	NP_221053	ATP-dependent protease ATP-binding subunit
1a	BupsA.00031.a	76809590	YP_334834	cyanate hydratase
1a	BrabA.00032.a	82700850	YP_415424	S-adenosyl-L-homocysteine hydrolase:ATP/GTP-binding site motif A (P-loop)
1a	BupsA.00032.a	76808996	YP_331474	S-adenosyl-L-homocysteine hydrolase
1a	BrabA.00033.a	83269666	YP_418957	Protein of unknown function DUF25
1a	BupsA.00033.a	76811000	YP_333853	methionine-R-sulfoxide reductase
1a	BrabA.00034.a	82699238	YP_413812	Alpha-beta hydrolase fold:Esterase/lipase/thioesterase, active site:Epoxide hydrolase:Prolyl aminopeptidase S33:Alpha/beta hy...
1a	BupsA.00035.a	76811410	YP_332167	ribose-phosphate pyrophosphokinase
1a	BrabA.00036.a	82699775	YP_414349	Acetyl-CoA carboxylase, biotin carboxylase:Carbamoyl-phosphate synthase L chain, ATP-binding:Carbamoyl-phosphate synthetase L...
1a	BupsA.00036.a	76811548	YP_332759	5-phosphoribosylglycinamide transformylase
1a	BrabA.00037.a	82699687	YP_414261	7Fe ferredoxin:3Fe-4S ferredoxin:4Fe-4S ferredoxin, iron-sulfur binding domain
1a	BupsA.00037.a	76809047	YP_332852	NADH dehydrogenase subunit I
1a	RiprA.00037.a	15604627	NP_221145	NADH dehydrogenase subunit I
1a	BrabA.00038.a	82699682	YP_414256	Pectinesterase:NADH-ubiquinone oxidoreductase, chain 49kDa
1a	BrabA.00039.a	82700109	YP_414683	Metallopeptidase family M24:Methionine aminopeptidase:Methionine aminopeptidase, subfamily 1
1a	BupsA.00039.a	76809214	YP_334031	methionine aminopeptidase, type I
1a	BupsA.00039.b	76809908	YP_333976	methionine aminopeptidase
1a	RiprA.00039.a	15604655	NP_221173	methionine aminopeptidase
1a	BupsA.00040.a	76811907	YP_331809	S-adenosylmethionine synthetase
1a	BrabA.00041.a	83269729	YP_419020	glutamate decarboxylase alpha
1a	BrabA.00042.a	82700669	YP_414643	Elongation factor, GTP-binding:ATP/GTP-binding site motif A (P-loop):Elongation factor Tu, C-terminal:Elongation factor Tu, d...
1a	BupsA.00043.a	76808528	YP_334796	dihydroorotase
1a	BrabA.00044.a	82699116	YP_413690	Zinc-containing alcohol dehydrogenase superfamily:Zinc-containing alcohol dehydrogenase
1a	RiprA.00045.a	15604567	NP_221085	DNA polymerase III subunit epsilon
1a	BupsA.00046.a	76809206	YP_332674	alpha.alpha-trehalose-phosphate synthase (UDP-forming)
1a	BupsA.00046.b	76810129	YP_334256	alpha.alpha-trehalose-phosphate synthase (UDP-forming)
1a	BrabA.00047.a	82700719	YP_415293	Fumarate lyase:Delta crystallin
1a	BupsA.00047.a	76810123	YP_334319	fumarate hydratase
1a	RiprA.00047.a	15604509	NP_221027	fumarate hydratase
1a	BupsA.00048.a	76809695	YP_334263	hypothetical protein BURPS1710b_2879
1a	BrabA.00049.a	83269711	YP_419002	Catalase
1a	BupsA.00050.a	76817894	YP_335917	GDP-mannose 4,6-dehydratase
1a	BupsA.00051.a	76808541	YP_331487	chemotaxis protein CheY
1a	BupsA.00051.b	76818365	YP_335540	response regulator
1a	BrabA.00052.a	82700517	YP_415091	Glyceraldehyde 3-phosphate dehydrogenase:TrkA potassium uptake protein:Glyceraldehyde-3-phosphate dehydrogenase, type 1
1a	BupsA.00052.a	76808621	YP_334837	glyceraldehyde-3-phosphate dehydrogenase, type 1
1a	BrabA.00053.a	82700834	YP_415408	CbxX/CfqX superfamily:ATP/GTP-binding site motif A (P-loop):AAA ATPase:Heat shock protein HslVU
1a	BupsA.00053.a	76812254	YP_331799	ATP-dependent protease ATP-binding subunit
1a	RiprA.00053.a	15604188	NP_220703	ATP-dependent protease ATP-binding subunit
1a	BupsA.00054.a	76809829	YP_334345	tRNA (guanine-N(1)-)-methyltransferase
1a	BupsA.00055.a	76809683	YP_334875	thiol peroxidase
1a	BupsA.00056.a	76810084	YP_332544	nicotinate-nucleotide pyrophosphorylase
1a	BrabA.00057.a	83269758	YP_419049	ATP/GTP-binding site motif A (P-loop):ABC transporter:AAA ATPase:TOBE domain
1a	BupsA.00057.a	76811324	YP_332452	ABC sugar transporter, ATP-binding protein
1a	BupsA.00058.a	76810139	YP_334452	thioredoxin-disulfide reductase
1a	BrabA.00059.a	82700034	YP_414608	Alkaline phosphatase
1a	BrabA.00060.a	82699088	YP_413662	3' exoribonuclease:Ribonuclease PH
1a	BupsA.00061.a	76818654	YP_336289	UDP-N-acetylglucosamine 2-epimerase
1a	BupsA.00062.a	76811886	YP_333941	inositol-5-monophosphate dehydrogenase
1a	BupsA.00063.a	76811640	YP_333748	histidyl-tRNA synthetase

1a	BupsA.00064.a	76810682	YP_332492	adenylate kinase
1a	BupsA.00065.a	76818936	YP_336708	4-hydroxybenzoate 3-monooxygenase
1a	BrabA.00066.a	82700043	YP_414617	RNA polymerase, alpha chain, bacterial and organelle
1a	BupsA.00066.a	76808919	YP_335114	DNA-directed RNA polymerase alpha subunit
1a	BrabA.00067.a	82700093	YP_414667	WW/Rsp5/WWP domain:Bacterial transferase hexapeptide repeat:Serine O-acetyltransferase
1a	BupsA.00068.a	76808615	YP_331622	phenylalanine-4-hydroxylase
1a	BrabA.00069.a	82700036	YP_414610	RecA bacterial DNA recombination protein:ATP/GTP-binding site motif A (P-loop):AAA ATPase
1a	BupsA.00069.a	76809737	YP_332384	recombinase A
1a	RiprA.00069.a	15604595	NP_221113	recombinase A
1a	BrabA.00070.a	83269089	YP_418380	AP endonuclease, family 2
1a	BrabA.00071.a	82700266	YP_414840	Methylenetetrahydrofolate reductase:5,10-methylenetetrahydrofolate reductase
1a	BupsA.00072.a	76811046	YP_332082	RNA methyltransferase, TrmH family, group 2
1a	BrabA.00073.a	82699840	YP_414414	Arsenate reductase:Arsenate reductase and related
1a	BupsA.00074.a	76817295	YP_336021	organic hydroperoxide resistance protein
1a	BupsA.00075.a	76809356	YP_335110	delta-aminolevulinic acid dehydratase
1a	RiprA.00075.a	15604396	NP_220912	delta-aminolevulinic acid dehydratase
1a	BrabA.00076.a	83269870	YP_419161	Phosphoribosyltransferase:Uracil phosphoribosyl transferase
1a	BupsA.00076.a	76811094	YP_332797	uracil phosphoribosyltransferase
1a	BupsA.00077.a	76811175	YP_334599	fumarylacetoacetase
1a	BupsA.00078.a	76809486	YP_331761	peptide deformylase
1a	BrabA.00079.a	82700648	YP_415222	Glutaredoxin:Thioredoxin type domain
1a	BrabA.00080.a	82700072	YP_414646	Ribosomal protein S12, bacterial and chloroplast forms:Ribosomal protein S12/S23
1a	RiprA.00080.a	15604007	NP_220522	30S ribosomal protein S12
1a	BupsA.00081.a	76810018	YP_334119	cysteine desulfurase IscS
1a	RiprA.00081.a	15604347	NP_220863	cysteine desulfurase
1a	BupsA.00082.a	76808885	YP_332640	porphobilinogen deaminase
1a	BrabA.00083.a	83269108	YP_418399	Cytochrome c heme-binding site:Cytochrome c, class IA/ IB:Cytochrome c, class I
1a	BupsA.00084.a	76811447	YP_334097	phosphopyruvate hydratase
1a	BrabA.00085.a	82699425	YP_413999	Short-chain dehydrogenase/reductase SDR:GDP-mannose 4,6-dehydratase
1a	BrabA.00085.b	82699910	YP_414484	NAD binding site:NAD-dependent epimerase/dehydratase:UDP-glucose 4-epimerase
1a	BupsA.00085.a	76808899	YP_334535	dTDP-glucose 4,6-dehydratase
1a	BupsA.00085.b	76809877	YP_334520	UDP-glucose 4-epimerase
1a	BupsA.00085.c	76812091	YP_334365	ADP-L-glycero-D-manno-heptose-6-epimerase
1a	BrabA.00086.a	82699777	YP_414351	Dehydroquinase, class II
1a	BupsA.00086.a	76809110	YP_334870	3-dehydroquininate dehydratase
1a	BrabA.00087.a	83269128	YP_418419	Chaperonin Cpn10
1a	BupsA.00088.a	76809529	YP_333522	ornithine carbamoyltransferase
1a	BrabA.00089.a	82700414	YP_414988	Bacterial periplasmic spermidine/putrescine-binding protein:Bacterial extracellular solute-binding protein, family 1
1a	BupsA.00090.a	76808557	YP_334707	3-deoxy-7-phosphoheptulonate synthase
1a	BupsA.00090.b	76819809	YP_335471	3-deoxy-7-phosphoheptulonate synthase
1a	BrabA.00091.a	82699919	YP_414493	GTP cyclohydrolase I
1a	BupsA.00092.a	76811221	YP_332521	isocitrate dehydrogenase
1a	BupsA.00093.a	76817422	YP_337624	N-ethylmaleimide reductase
1a	BupsA.00094.a	76809923	YP_331794	acetylglutamate kinase
1a	BupsA.00095.a	76819100	YP_337143	non-heme chloroperoxidase
1a	BupsA.00096.a	76819597	YP_336667	2-amino-3-ketobutyrate coenzyme A ligase
1a	BrabA.00097.a	82699252	YP_413826	Pantoate-beta-alanine ligase:Cytidyltransferase-related domain
1a	BupsA.00098.a	76809622	YP_333787	molybdopterin converting factor, subunit 2
1a	RiprA.00099.a	15604221	NP_220737	ATP synthase subunit E
1a	BupsA.00100.a	76810823	YP_334856	L-arabinose ABC transporter, periplasmic L-arabinose-binding protein
1a	BupsA.00101.a	76812239	YP_334027	ferredoxin--NADP reductase
1a	BrabA.00102.a	82699973	YP_414547	DAHPh synthetase I/KDSA superfamily:2-dehydro-3-deoxyphosphoacetate aldolase
1a	BupsA.00102.a	76811601	YP_334635	2-dehydro-3-deoxyphosphoacetate aldolase
1a	RiprA.00102.a	15603941	NP_220456	2-dehydro-3-deoxyphosphoacetate aldolase
1a	BupsA.00103.a	76811233	YP_334945	uracil-DNA glycosylase
1a	BupsA.00104.a	76819356	YP_337703	beta-lactamase
1a	BupsA.00105.a	76819170	YP_335488	beta-lactamase
1a	BrabA.00106.a	83269117	YP_418408	Methyladenine glycosylase
1a	BupsA.00107.a	76810945	YP_332430	efflux transporter, RND family, MFP subunit subfamily
1a	BupsA.00108.a	76808910	YP_332194	single-strand DNA-binding protein
1a	BupsA.00109.a	76818168	YP_337486	cytosine deaminase
1a	BupsA.00110.a	76808994	YP_334545	Rubredoxin
1a	BrabA.00111.a	82699855	YP_414429	Glutamine synthetase class-I, adenylation site:Glutamine synthetase type I:Glutamine synthetase, catalytic domain:Glutamine s...
1a	BupsA.00112.a	76808721	YP_334688	3-methyl-2-oxobutanoate hydroxymethyltransferase
1a	BrabA.00113.a	82700919	YP_415493	Beta-ketoacyl synthase
1a	BupsA.00114.a	76811388	YP_332076	phosphoglyceromutase
1a	BupsA.00115.a	76810731	YP_332531	deoxyuridine 5'-triphosphate nucleotidohydrolase
1a	BupsA.00116.a	76809167	YP_332286	4-hydroxythreonine-4-phosphate dehydrogenase
1a	BupsA.00117.a	76818492	YP_336145	catechol 1,2-dioxygenase
1a	BupsA.00118.a	76811752	YP_334534	glucose-1-phosphate thymidyltransferase
1a	BupsA.00119.a	76810239	YP_334912	D-alanine--D-alanine ligase B
1a	BupsA.00120.a	76812026	YP_334893	GTP-binding protein, GTP1/OBG family
1a	BrabA.00121.a	83269625	YP_418916	Biotin synthase:Elongator protein 3/MiaB/NifB:Radical SAM
1a	BupsA.00122.a	76808522	YP_333903	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
1a	BrabA.00123.a	83269912	YP_419203	Immunoglobulin/major histocompatibility complex
1a	BupsA.00124.a	76811450	YP_335055	imidazoleglycerol-phosphate dehydratase
1a	BupsA.00125.a	76812213	YP_331490	chemotaxis protein methyltransferase
1a	BupsA.00126.a	76809833	YP_332614	aspartate 1-decarboxylase precursor
1a	BupsA.00127.a	76808819	YP_333607	NIPSNAP superfamily
1a	BrabA.00128.a	83269728	YP_419019	Bacterial extracellular solute-binding protein, family 3:Pyridoxal-dependent decarboxylase
1a	BupsA.00129.a	76811002	YP_334247	pyridoxal kinase
1a	BupsA.00130.a	76817652	YP_336065	peptidyl-prolyl cis-trans isomerase
1a	BupsA.00131.a	76809280	YP_334273	pyridoxal phosphate biosynthetic protein
1a	RiprA.00132.a	15604168	NP_220683	ribose-5-phosphate isomerase B
1a	BupsA.00133.a	76809388	YP_334067	cysteinyl-tRNA synthetase
1a	BupsA.00134.a	76811849	YP_331434	4-hydroxyphenylpyruvate dioxygenase
1a	RiprA.00135.a	15604558	NP_221076	queuine tRNA-ribosyltransferase
1a	BrabA.00136.a	82700896	YP_415470	Ferritin:DNA-binding protein Dps
1a	BupsA.00137.a	76811379	YP_335115	30S ribosomal protein S4
1b	BrabB.00011.a	82700242	YP_414816	cell division protein FtsZ
1b	BrabB.00011.a	62290318	YP_222111	cell division protein FtsZ
1b	BrmeA.00011.a	17986868	NP_539502	cell division protein FtsZ
1b	BrsuA.00011.a	23502296	NP_698423	cell division protein FtsZ
1b	BumaA.00011.a	53726048	YP_104088	cell division protein FtsZ
1b	BupsA.00011.a	76810415	YP_334909	cell division protein FtsZ
1b	BupsB.00011.a	157938681	EDO94351	cell division protein FtsZ
1b	ButhA.00011.a	83719008	YP_441671	cell division protein FtsZ
1b	RicoA.00011.a	15892938	NP_360652	cell division protein FtsZ
1b	RifeA.00011.a	67458662	YP_246286	cell division protein FtsZ
1b	RiprA.00011.a	15604510	NP_221028	cell division protein FtsZ
1b	RiriA.00011.a	157828869	YP_001495111	cell division protein FtsZ
1b	RityA.00011.a	51473844	YP_067601	cell division protein FtsZ

Table 3 – Species key

Organism	Code
<i>Brucella abortus</i> biovar 1 str. 9-941	BrabB
<i>Brucella melitensis</i> 16M	BrmeA
<i>Brucella melitensis</i> biovar Abortus 2308	BrabA
<i>Brucella suis</i> 1330	BrsuA
<i>Burkholderia mallei</i> ATCC 23344	BumaA
<i>Burkholderia pseudomallei</i> 1710b	BupsA
<i>Burkholderia pseudomallei</i> Pasteur 52237	BupsB
<i>Burkholderia thailandensis</i> E264	ButhA
<i>Rickettsia conorii</i> str. Malish 7	RicoA
<i>Rickettsia felis</i> URRWXCal2	RifeA
<i>Rickettsia prowazekii</i> str. Madrid E	RiprA
<i>Rickettsia rickettsii</i> str. Sheila Smith	RiriA
<i>Rickettsia typhi</i> str. Wilmington	RityA